ORIGINAL

Desarrollo de una escala predictiva en pacientes con *shock* séptico refractario mediante un modelo híbrido de aprendizaje automático y regresión

Sejin Heo*[1], Daun Jeong*[2,3], Minyoung Choi[1], Inkyu Kim[4], Minha Kim[1], Ye Rim Lee[4], Byuk Sung Ko[5], Seung Mok Ryoo[6], Eunah Han[7], Hyunglan Chang[8], Chang June Yune[9], Hui Jai Lee[10,11], Gil Joon Suh[11], Sung-Hyuk Choi[12], Sung Phil Chung[7], Tae Ho Lim[5], Won Young Kim[6], Kyuseok Kim[8], Sung Yeon Hwang[1], Jong Eun Park[1], Gun Tak Lee[1], Tae Gun Shin[1,4] en nombre de la Korean Shock Society

Objetivos. Desarrollar una escala predictiva en pacientes con *shock* séptico (SS) refractario basada en parámetros clínicos disponibles en la valoración inicial.

Métodos. Estudio multicéntrico y retrospectivo que analizó pacientes de la base MIMIC-IV *(Marketplace for Medical Information in Intensive Care)* para la derivación y validación interna de la escala. Se realizó una validación externa con pacientes con SS diagnosticados en el servicio de urgencias (cohorte SU) y con pacientes diagnosticados en la unidad de cuidados intensivos (cohorte UCI). La escala de SS refractario (ESSR) se desarrolló con *AutoScore.* El rendimiento se analizó mediante el área bajo la curva (ABC) de la curva operativa del receptor (ROC). El resultado primario fue el desarrollo de SS refractario en las 24 horas siguientes al ingreso en la UCI o de la llegada al SU, definido como la necesidad de una dosis equivalente de noradrenalina superior a 0,5 µg/kg/min.

Resultados. Se analizaron 29.618 pacientes de la base MIMIC-IV, 3.113 de la cohorte SU y 1.015 de la cohorte UCI. La ESSR incluyó seis variables: lactato, presión arterial sistólica, frecuencia cardiaca, temperatura, pH arterial y cifra de leucocitos. El ABC ROC fue de 0,873 (IC 95%: 0,846-0,900) en la validación interna; 0,705 (IC 95% 0,678-0,733) en la cohorte SU a su llegada y 0,781 (IC 95%: 0,757-0,805) en el momento de la hipoperfusión o hipotensión; y 0,822 (IC 95%: 0,787-0,857) en la cohorte UCI. La calibración fue aceptable en todas las cohortes.

Conclusiones. La ESSR tuvo una precisión diagnóstica adecuada en múltiples cohortes de validación de pacientes incluidos en el SU y en la UCI.

Palabras clave: Shock séptico. Vasopresores. Predictivo. Aprendizaje automático.

Scale to predict risk for refractory septic shock based on a hybrid approach using machine learning and regression modeling

Objective. To develop a scale to predict refractory septic shock (SS) based on clinical variables recorded during initial evaluations of patients.

Methods. Multicenter retrospective study of data for patients with suspected infection registered in the Marketplace for Medical Information in Intensive Care (MIMIC-IV). These data were used for the development and internal validation of the refractory SS scale (RSSS). For external validation, we used retrospective data for 2 cohorts: 1) patients diagnosed with SS in an emergency department (ED cohort) whose data were registered in a Korean SS registry, and 2) patients diagnosed with SS in 6 hospital intensive care units (ICU cohort). A machine-learning automatic clinical scoring system (AutoScore) was used in the development phase. The performance of the RSSS in the validation cohorts was assessed with the area under the receiver operating characteristic curve (AUROC) for each. The primary outcome was the development of refractory SS within 24 hours of ICU admission. Refractory SS was defined by the need for a norepinephrine-equivalent dose greater than 0.5 µg/kg/min.

Results. We collected data for 29 618 patients from the MIMIC-IV registry, 3113 patients for the ED cohort, and 1015 for the ICU cohort. The RSSS had 6 predictors: serum lactate level, systolic blood pressure, heart rate, temperature, arterial pH, and leukocyte count. The scale's AUROCs were as follows: 0.873 (95% CI, 0.846-0.900) in the internal validation, 0.705 (95% CI, 0.678-0.733) in the ED cohort on arrival, 0.781 (95% CI, 0.757-0.805) in the ED cohort at the moment of diagnosing hypoperfusion or hypotension, and 0.822 (95% CI, 0.787-0.857) in the ICU cohort. Calibration was acceptable in all the cohorts.

Conclusion. The RSSS had adequate diagnostic accuracy in multiple cohorts of patients diagnosed in the ED and ICU.

Keywords: Septic shock. Vasopressors. Risk. Machine learning.

DOI: 10.55633/s3me/108.2024

*Ambos autores han contribuido por igual en este trabajo.

Filiación de los autores:
[1]Department of Emergency Medicine, Samsung Medical Centre, Sungkyunkwan University School of Medicine, Seoul, Korea.
[2]Division of Critical Care Medicine, Department of Emergency Medicine, Chung-Ang University Gwangmyeong Hospital, Gwangmyeong-si, Gyeonggi-do, Republic of Korea.
[3]Department of Emergency Medicine, Chung-Ang University College of Medicine, Seoul, Korea.
[4]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Korea.
[5]Department of Emergency Medicine, College of Medicine, Hanyang University, Seoul, Korea.
[6]Department of Emergency Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea.
[7]Department of Emergency Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea.
[8]Department of Emergency Medicine, CHA Bundang Medical Center, CHA University School of Medicine, Seongnam, Korea.
[9]Department of Critical Care Medicine, Anyang Sam Hospital, An Yang, Korea.
[10]Department of Emergency Medicine, SMG-SNU Boramae Medical Center, Seoul, Korea.
[11]Department of Emergency Medicine, Seoul National University College of Medicine, Seoul, Korea.
[12]Department of Emergency Medicine, College of Medicine, Korea University, Seoul, Korea.

Contribución de los autores:
Todos los autores han confirmado su autoría en el documento de responsabilidades del autor, acuerdo de publicación y cesión de derechos a EMERGENCIAS.

Autor para correspondencia:
Tae Gun Shin, Gun Tak Lee. Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea.

Correo electrónico:
taegunshin@skku.edu, actionamigo@gmail.com

Información del artículo:
Recibido: 5-4-2022
Aceptado: 8-8-2024
Online: 3-12-2024

Editor responsable:
Aitor Alquézar Arbé

DOI:
10.55633/s3me/108.2024

SHOUT
SHARE
IT

Introducción

El *shock* séptico (SS) es la complicación más grave de la sepsis, que implica una mortalidad alta, entre el 25 y el 50%, y se caracteriza por disfunción circulatoria, celular y metabólica[1]. Para mejorar el pronóstico, es necesario identificar esta entidad de forma precoz e iniciar el tratamiento apropiado[1-3]. La presentación clínica del SS es variada, con diferentes respuestas hemodinámicas, lo que dificulta su manejo[4,5]. El *shock* séptico refractario (SSR) representa la forma más grave y conlleva una mortalidad superior al 50%. Aunque no existe una definición de consenso, el SSR requiere un abordaje multidisciplinar y se caracteriza por no responder al tratamiento habitual y precisar altas dosis de noradrenalina (NA)[6-8].

Predecir de forma temprana la progresión a SSR facilitaría el inicio de medidas enérgicas como el soporte hemodinámico avanzado, el uso de vasopresores y plantear estrategias de soporte de órganos. Según las recomendaciones de la *Surviving Sepsis Campaign* (SSC), la administración de dosis más altas de NA se asocia a la decisión de utilizar un segundo vasopresor y administrar corticosteroides[9].

El objetivo del estudio es desarrollar y validar una nueva escala predictiva –escala de *shock* séptico refractario (ESSR)– basada en *AutoScore,* una técnica híbrida de aprendizaje automático *(machine learning)* y regresión[10-13]. Se analizan variables clínicas disponibles durante la valoración inicial en los servicios de urgencias (SU) y en las unidades de cuidados intensivos (UCI).

Método

Diseño del estudio

Se trata de un estudio multicéntrico y retrospectivo, diseñado para desarrollar una escala predictiva de SSR en pacientes atendidos en el SU o que ingresaron en UCI. Se utilizaron tres fuentes de datos: la base MIMIC-IV *(Marketplace for Medical Information in Intensive Care)* para la derivación y validación interna, y una cohorte retrospectiva de pacientes ingresados en UCI y una cohorte prospectiva de pacientes incluidos en el SU para la validaciones externas.

La base MIMIC-IV contiene información clínica completa y anonimizada de pacientes que acudieron al SU o fueron ingresados en la UCI del Beth Israel Deaconess Medical Center, un hospital universitario de Boston (Massachusetts, Estados Unidos), entre 2008 y 2019[14]. Se incluyeron pacientes de edad ≥ 18 años con sospecha de infección en el primer día de ingreso en la UCI de acuerdo al uso de antibiótico y a la obtención de cultivos microbiológicos. Información adicional sobre los pacientes de esta cohorte aparece en una publicación previa[15]. La base MIMIC-IV se dividió de forma aleatoria en un conjunto de derivación (70%), uno de validación (10%) y uno de prueba (20%).

Para la validación externa, se utilizaron dos cohortes. Por una parte, una cohorte de pacientes con SS –registro de la *Korean Shock Society*– reclutados en el SU desde octubre de 2015 hasta diciembre de 2019. Esta cohorte incluyó pacientes de edad ≥ 18 años con infección confirmada o con alta sospecha e hipotensión refractaria, sin respuesta a la reanimación inicial con volumen o hipoperfusión, y lactato sérico ≥ 4 mmol/L. Información detallada sobre esta cohorte se describe en publicaciones previas[16,17]. Por otra parte, se utilizó una cohorte de pacientes con sepsis –según los criterios Sepsis 31–, ingresados en UCI de 6 hospitales coreanos. Se incluyeron pacientes ingresados en UCI médica o quirúrgica y con sepsis[18].

Recogida de datos

Tras cumplir el acuerdo de uso de datos y completar la formación *Protection of Human Subjects,* los investigadores accedieron a la base MIMIC-IV a través del enlace https://mimic.physionet.org y completaron la *Collaborative Institutional Training Initiative.* A continuación, se utilizó PostgreSQL versión 10.16 (PostgreSQL Global Development Group, Berkeley, California, EE.UU.) y Python versión 3.7 para la extracción de las siguientes variables: edad, sexo, constantes vitales, analíticas del primer día, uso de fármacos vasoactivos, puntuación de la escala SOFA *(Sequential Organ-Failure Assessment)* y mortalidad intrahospitalaria[19]. Los códigos de datos procedían del *MIMIC Code Repository*[20]. Sólo se identificó la sospecha del foco de infección.

La cohorte SU recogió los datos de forma prospectiva mediante un registro informatizado. La cohorte UCI recogió la información de forma retrospectiva. Un comité de calidad supervisó la integridad y consistencia de los datos e informó a los investigadores de las incongruencias. Los datos de supervivencia se extrajeron del propio registro, así como de la historia clínica de los hospitales y del registro nacional de mortalidad.

Variables

La variable de resultado principal fue el desarrollo de SSR en las 24 horas siguientes al ingreso en la UCI o de la llegada al SU. El SSR se definió por unos requerimientos de dosis equivalente de NA ≥ 0,5 µg/kg/min[8,21,22]. Cada dosis de vasopresor (dopamina, adrenalina, fenilefrina o vasopresina) se convirtió en una dosis equivalente de NA (https://revistaemergencias.org/extra/MS-4528.pdf Tabla S1)[23]. Se utilizaron las dosis máximas administradas durante al menos una hora en las primeras 24 horas desde la llegada al SU o del ingreso en UCI. Otras variables de resultado fueron la mortalidad intrahospitalaria y la dosis de NA.

Para diseñar la escala se evaluaron 17 variables independientes: edad, presión arterial sistólica, presión arterial diastólica, presión arterial media, frecuencia cardiaca, frecuencia respiratoria, temperatura, recuento de leucocitos, hemoglobina, recuento de plaquetas, nitrógeno ureico, creatinina, bilirrubina total, lactato, pH arterial, presión parcial de oxígeno arterial y presión parcial de dióxido de carbono arterial.

Generación de la escala con AutoScore

Para la creación de la escala se utilizó *AutoScore.* Este programa utiliza aprendizaje automático y regresión y tiene seis módulos que permiten la selección de variables, la generación de escalas, su ajuste y validación. En el módulo 1, se aplicó la técnica *random forest* para clasificar todas las variables estudiadas en función de su importancia. El número de árboles de decisión del *random forest* fue de 100. Se utilizaron los cuantiles para el ajuste de cada variable.

Se generaron gráficos de parsimonia de las áreas bajo la curva (ABC) de la curva operativa del receptor (ROC) para determinar las variables más relevantes. Se realizaron *splines* cúbicos restringidos para evaluar la relación entre las variables predictoras y la variable de resultado.

El módulo 2 de *AutoScore* se utilizó para categorizar las variables continuas con el fin de simplificar la interpretación. Los módulos 3 y 4 se utilizaron para desarrollar la puntuación de la escala mediante regresión logística y seleccionar las variables del modelo. Con el Módulo 5 y en base a trabajos previos y al contexto clínico, se completó la selección de variables y se ajustaron los intervalos de las variables continuas. El Módulo 6 evaluó el rendimiento de la escala final. Puede encontrarse información detallada sobre el programa *AutoScore* en publicaciones previas[10,24].

Evaluación de la escala

Se analizó el rendimiento del modelo con el cálculo del ABC de la ROC con remuestreos por *bootstrapping.* También se evaluaron los datos de la base MIMIC-IV utilizando un muestreo aleatorio múltiple. Se hicieron gráficos de calibración. En función de sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN) y la incidencia, se obtuvieron dos puntos de decisión que generaron tres categorías de riesgo (bajo, intermedio y alto). Estos parámetros y las ABC de la ROC se expresan con un intervalo de confianza del 95% (IC 95%). A continuación, se analizó la frecuencia de SSR, los pacientes que precisaron dosis de NA $\geq$ 0,25 µg/kg/min y la mortalidad intrahospitalaria en cada grupo de riesgo.

Se realizó una validación interna y también una validación externa con las cohortes procedentes de UCI y SU. Con la cohorte SU se usaron dos momentos temporales: a la llegada al SU y en el momento del diagnóstico de la hipoperfusión o de la hipotensión. Los valores de lactato iniciales se utilizaron para calcular la puntuación en el momento inicial. Los valores de lactato tras la administración de fluidos se utilizaron para calcular la puntuación en el momento de la hipoperfusión o hipotensión. El modelo final se eligió en función de su capacidad predictiva en la cohorte de prueba. Elegimos un modelo final basándonos en su rendimiento en la predicción del SSR en el conjunto de datos de prueba. Posteriormente, se realizó un análisis de sensibilidad utilizando datos completos sin valores perdidos en las cohortes de validación. Además, se probaron otros modelos utilizando múltiples aproximaciones estadísticas, como *random forest* y regresión logística por pasos.

Análisis estadístico

Las variables categóricas se muestran como frecuencias y porcentajes, y se compararon con la prueba ji cuadrado. Las variables continuas se presentan como medias y desviaciones estándar (DE), y se compararon mediante la prueba t de Student. Un valor de p bilateral < 0,05 se consideró estadísticamente significativo. Los valores extremos e inviables se consideraron ausentes. Los valores perdidos de las variables continuas se imputaron con valores medios. El desarrollo del modelo predictivo se describe en el apartado previo. En la cohorte SU se realizó una regresión logística para evaluar la asociación entre la mortalidad intrahospitalaria y el intervalo de tiempo entre la administración del primer y del segundo vasopresor en cada grupo de riesgo. Todos los análisis se realizaron utilizando el programa R versión 3.6.3 (R Foundation for Statistical Computing, Viena, Austria).

Aspectos éticos

El estudio fue aprobado por los diferentes hospitales participantes: Samsung Medical Center (No. 2023-08-146), Seoul Metropolitan Government–Seoul National University Boramae Medical Center (No. 30-2021-139), Asan Medical Center (No. 2022-0203), Gangnam Severance Hospital (No. 3-2021-0506), Hanyang University Medical Center (No. 2022-01-019), Anyang Sam Hospital (No. 2022007), and CHA Bundang Medical Center (No.2022-07-070).

Resultados

Características de los pacientes

La base MIMIC-IV incluyó 69.211, de los que se analizaron 29.618 pacientes y hubo 995 (3,4%) con SSR (Figura 1). Se asignaron 20.715 pacientes (70%) a la cohorte de derivación y 8.903 (30%) a las cohortes de validación y prueba. La edad media fue de 65 años y el 56,7% eran varones. La mortalidad intrahospitalaria fue del 12,4% y la media de la escala SOFA fue de 6 puntos. La dosis media de NA fue de 0,07 µg/kg/min.

De los 4.180 pacientes de la cohorte SU, se analizaron 3.113 pacientes y 413 (9,9%) presentaron un SSR. Los pacientes tenían una media de 67 años y el 40,7% eran varones. La mortalidad intrahospitalaria fue del 20,9% y la media de la escala SOFA fue de 6 puntos. La dosis media de NA fue de 0,25 µg/kg/min.

La cohorte UCI incluyó 1.015 pacientes, 168 (16,6%) con SSR. La edad media fue de 71 años y el 42,7% eran varones. La mortalidad intrahospitalaria fue del 37,6% y la media de la escala SOFA fue de 9 puntos. La dosis media de NA fue de 0,27 µg/kg/min (1,01). Las características y evolución de los pacientes se muestran en la Tabla 1.

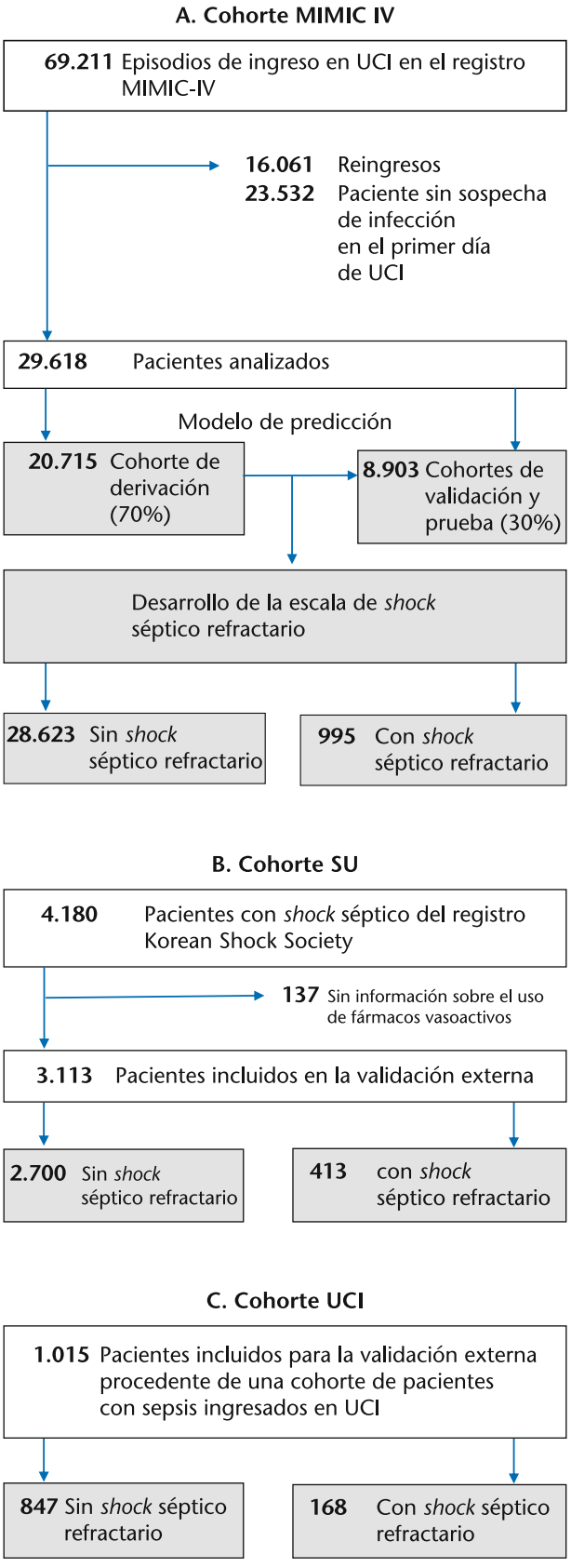
A. Cohorte MIMIC IV
69.211 Episodios de ingreso en UCI en el registro MIMIC-IV
16.061 Reingresos
23.532 Paciente sin sospecha de infección en el primer día de UCI
29.618 Pacientes analizados
Modelo de predicción
20.715 Cohorte de derivación (70%)
8.903 Cohortes de validación y prueba (30%)
Desarrollo de la escala de shock séptico refractario
28.623 Sin shock séptico refractario
995 Con shock séptico refractario
B. Cohorte SU
4.180 Pacientes con shock séptico del registro Korean Shock Society
137 Sin información sobre el uso de fármacos vasoactivos
3.113 Pacientes incluidos en la validación externa
2.700 Sin shock séptico refractario
413 con shock séptico refractario
C. Cohorte UCI
1.015 Pacientes incluidos para la validación externa procedente de una cohorte de pacientes con sepsis ingresados en UCI
847 Sin shock séptico refractario
168 Con shock séptico refractario

Figura 1. Diagrama de flujo de inclusion de pacientes en las tres cohortes (A, B y C) utilizadas en este estudio.
UCI: unidad de cuidados intensivos; SU: servicio de urgencias.

Tabla 1. Características basales de las cohortes de derivación y validación

	Cohorte MIMIC-IV N = 29.618 n (%)	Cohorte Urgencias N = 3.113 n (%)	Cohorte UCI N = 1.015 n (%)
Edad (años) [media (DE)]	65 (16)	67 (13)	71 (14)
Sexo Varón	16.806 (56,7)	1.268 (40,7)	433 (42,7)
Frecuencia cardiaca (lpm) [media (DE)]	89,8 (20,1)	109,8 (25,6)	100,4 (24,5)
Frecuencia respiratoria (rpm) [media (DE)]	19,2 (6,3)	21,3 (5,1)	22,1 (6,7)
PAS (mmHg) [media (DE)]	122,4 (24,9)	100,3 (29,1)	116,9 (28,1)
PAD (mmHg) [media (DE)]	65,9 (17,2)	60,5 (18,7)	68,1 (17,8)
Temperatura [media (DE)]	36,7 (0,9)	37,8 (1,3)	36,9 (1,1)
Leucocitos (10^3/L) [media (DE)]	13,1 (10,9)	12,5 (17,9)	12,2 (25,7)
Hemoglobina (g/dL) [media (DE)]	11,8 (2,3)	10,8 (2,5)	10,5 (4,7)
Urea (mg/dL) [media (DE)]	26,4 (21,9)	32,2 (21,6)	36,5 (26,1)
Creatinina (mg/dL) [media (DE)]	1,4 (1,5)	1,8 (1,6)	2,1 (7,7)
Plaquetas(10^3/L) [media (DE)]	197,9 (86,5)	159,1 (125,1)	96,5 (121,0)
Lactato (mmol/L) [media (DE)]	2,36 (2,17)	4,28 (3,08)	4,21 (20,06)
pH arterial [media (DE)]	7,36 (0,10)	7,42 (0,11)	7,38 (0,12)
PaO_2 (mmHg) [media (DE)]	170,9 (122,0)	87,4 (42,0)	103,4 (54,7)
$PaCO_2$ (mmHg) [media (DE)]	42,8 (12,9)	29,8 (13,2)	35,3 (16,7)
Dopamina	601 (2,0)	46 (1,5)	73 (7,2)
Noradrenalina	5,392 (18,2)	2,755 (88,5)	609 (60,0)
Adrenalina	1,489 (5,0)	187 (6,0)	61 (6,0)
Vasopresina	1,611 (5,4)	809 (26,0)	137 (13,5)
Dosis equivalente de noradrenalina (µg/kg/min) [media (DE)]	0,07 (0,18)	0,25 (0,47)	0,27 (1,01)
Terapia substitución renal	629 (2,1)	411 (13,2)	215 (21,2)
Ventilación mecánica	14.812 (50,0)	2.209 (71,0)	553 (54,6)
Escala SOFA (puntos) [media (DE)]	5,5 (3,8)	6,2 (3,2)	8,8 (4,1)
Shock séptico (Sepsis-3)	6.501 (22,0)	2.594 (83,3)	412 (40,6)
Shock séptico refractario	995 (3,4)	413 (13,3)	168 (16,6)
Mortalidad intrahospitalaria	3.675 (12,4)	650 (20,9)	382 (37,6)

DE: desviación estándar; lpm: latidos por minuto; PAS: presión arterial sistólica; PAD: presión arterial diastólica; rpm: respiraciones por minuto; SOFA: *sequential organ failure assessment;* UCI: unidad de cuidados intensivos.

Generación de la Escala de Shock *Séptico Refractario*

Tras el desarrollo de la escala se seleccionaron seis variables: temperatura, frecuencia cardiaca, presión arterial sistólica, lactato, pH arterial y recuento de leucocitos. Estas variables se seleccionaron en base a sus características y a los gráficos de parsimonia (Tabla 2, Figura 2). Se utilizaron curvas *spline* restringidas para mostrar la relación entre los diferentes valores de cada variable y la probabilidad de SSR (https://revistaemergencias.org/extra/MS-4528.pdf Figura S1). La incidencia de SSR aumentó a medida que aumentaba la puntuación ESSR en todos los subgrupos de datos utilizados y la escala

Tabla 2. Puntuación de la escala para predecir *shock* séptico refractario

Variable e intervalo	Puntos
Lactato (mmol/L)	
< 2	0
2-4	3
4-6	6
≥ 6	7
PAS (mmHg)	
< 80	8
80-100	5
100-140	2
≥ 140	0
Leucocitos (10^3/L)	
< 4	3
4-25	0
≥ 25	1
pH arterial	
< 7,2	6
7,2-7,35	4
≥ 7,35	0
Temperatura (°C)	
< 35	3
35-36	2
36-38	0
≥ 38	1
Frecuencia cardiaca (lpm)	
< 60	1
60-100	0
100-120	1
≥ 120	3

PAS: presión arterial sistólica; lpm: latidos por minuto.

mostró una calibración adecuada (https://revistaemergencias.org/extra/MS-4528.pdf Figura S2).

Capacidad predictiva de la Escala de Shock *Séptico Refractario*

El ABC ROC de la ESSR para predecir SSR fue de 0,873 (IC 95%: 0,846-0,900) en la validación interna (https://revistaemergencias.org/extra/MS-4528.pdf Figura S3) y de 0,895 (IC 95%: 0,870-0,920) en toda la cohorte MIMIC-IV con muestreo aleatorio (https://revistaemergencias.org/extra/MS-4528.pdf Tabla S2). La comparación del rendimiento entre la puntuación ESSR y otros modelos en la validación interna con MIMIC-IV se muestra en la https://revistaemergencias.org/extra/MS-4528.pdf Tabla S2. El modelo *random forest* con las 17 variables mostró el ABC de la ROC más alto (0,899; IC 95%: 0,876-0,923), pero el modelo final de la ESSR –con 7 variables– fue comparable o superior a otros modelos, incluido el *random forest* con 6 variables (ABC ROC 0,866), el modelo de regresión logística con 17 variables (ABC ROC 0,865) y el modelo por pasos con 14 variables (ABC ROC 0,864).

En la cohorte SU, el ABC ROC fue de 0,705 (IC 95%: 0,678-0,733) a la llegada a urgencias y de 0,781 (IC 95%: 0,757-0,805) en el momento del diagnóstico de hipoperfusión o hipotensión. En la cohorte UCI, el ABC ROC fue de 0,822 (IC 95% 0,787-0,857) (https://revistaemergencias.org/extra/MS-4528.pdf Figura S3). En el análisis de sensibilidad con los conjuntos de datos completos, los resultados fueron coherentes con los de los análisis principales (https://revistaemergencias.org/extra/MS-4528.pdf Tabla S4).

Estratificación del riesgo con la Escala de Shock *Séptico Refractario*

Los grupos de bajo riesgo (ESSR < 6 puntos) y alto riesgo (ESSR ≥ 12 puntos) se seleccionaron con un VPP del 5% (sensibilidad del 90%) y del 20% (especificidad del 90%) en los datos de MIMIC-IV. La distribución del número de pacientes y la incidencia de SSR según el grupo de riesgo ESSR se muestran en la https://revistaemergencias.org/extra/MS-4528.pdf Figura S3. La sensibilidad y el VPN se mantuvieron altos (90% a 98%) y el VPP osciló entre el 7% y el 14% con el punto de decisión de bajo riesgo. En el grupo de alto riesgo, el VPP osciló entre el 20 y el 40% y la especificidad entre el 53 y el 90% según la cohorte. El rendimiento detallado de la estratificación del riesgo de la ESSR en todas las cohortes se muestra en la Tabla 3.

En los pacientes de la base MIMIC-IV, la incidencia de SSR fue del 0,5% en el grupo de bajo riesgo, del 2,5% en el grupo de riesgo intermedio y del 20,1% en el grupo de alto riesgo (Tabla 4). La mortalidad intrahospitalaria fue del 6,1%, 13,7% y 37,5% en los grupos de bajo, intermedio y alto riesgo, respectivamente. En la cohorte SU, la incidencia de SSR aumentó del 2% al 19,5% y la mortalidad intrahospitalaria aumentó del 12,1% al 26,9% según la estratificación de riesgo de la ESSR a la llegada a urgencias. Se observaron tendencias similares según la puntuación ESSR en el momento del diagnóstico de hipoperfusión o hipotensión en la cohorte SU, así como en la cohorte UCI. La proporción de pacientes que requirieron una dosis de NA ≥ 0,25 µg/kg/min también mostró tendencias crecientes en todas las cohortes.

Tiempo hasta la administración del segundo vasopresor y mortalidad intrahospitalaria según la puntuación de la Escala de Shock *Séptico Refractario en el SU*

En la cohorte SU, se evaluó la asociación entre el tiempo transcurrido desde la administración del primer vasopresor hasta la del segundo y la mortalidad (https://revistaemergencias.org/extra/MS-4528.pdf Tabla S4). Este intervalo se asoció significativamente con la mortalidad intrahospitalaria en el grupo de alto riesgo: *Odds Ratio* (OR) no ajustada 1,07 (IC 95%: 1,03-1,12); OR ajustada 1,09 (IC 95%: 1,04-1,14). No se observó ninguna asociación estadísticamente significativa en los grupos de riesgo bajo o intermedio.

Discusión

Este estudio utilizó un método híbrido de aprendizaje automático-regresión mediante *AutoScore,* en una base

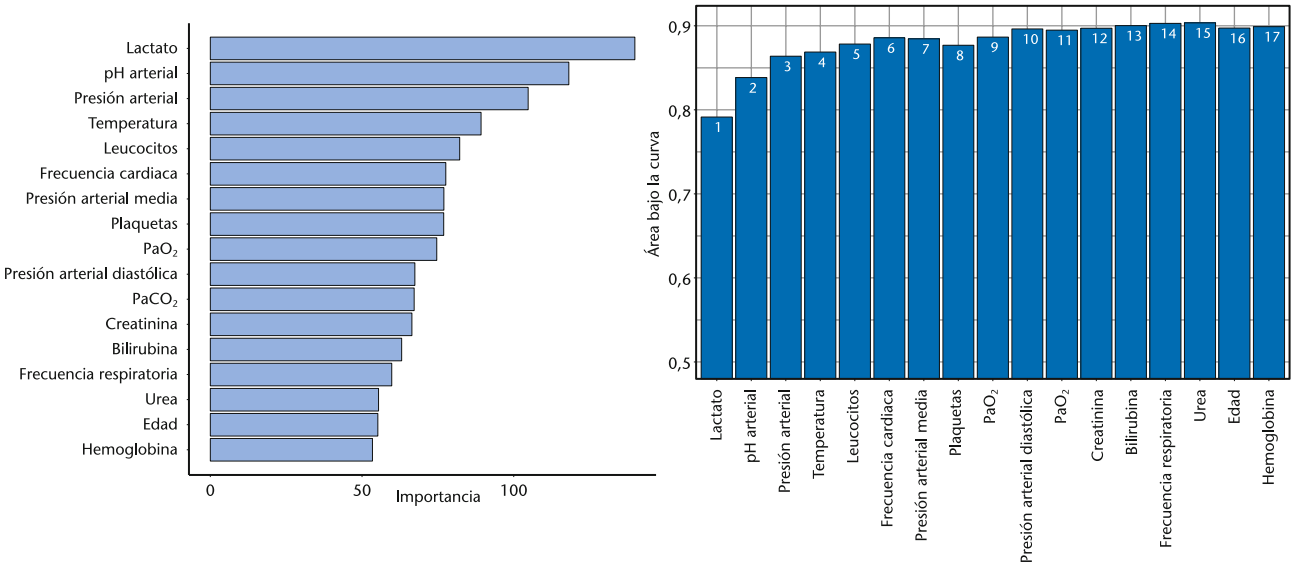

Figura 2. Características relevantes (izquierda) y gráfico de parsimonia (derecha) de los datos del registro MIMIC-IV.

de datos de UCI abierta para desarrollar una escala predictiva de SSR. En nuestro conocimiento, la ESSR es la primera escala diseñada para predecir SSR de forma precoz en el SU y en UCI. Se ha realizado la validación externa en una cohorte UCI y en una cohorte SU. La escala tuvo una buena capacidad predictiva con implicaciones potenciales como permitir identificar de forma precoz a los pacientes con SS de alto riesgo.

El SSR es una complicación grave de la sepsis y se caracteriza por no responder al tratamiento inicial. Los pacientes con SSR se han definido de forma heterogénea: presencia de hipotensión, disfunción de órganos, requerimientos de apoyo vasopresor a dosis altas, a menudo superior a 0,5 µg/kg/min de NA, y que presentan un rápido empeoramiento clínico[6-8,16,25]. Cabe resaltar que los pacientes que requieren NA ≥ 0,5 µg/kg/min presentan una mortalidad asociada de hasta el 60%, cifra consistente con los resultados del presente estudio[7,8,16].

El retraso en el inicio de vasopresina en pacientes con altos requerimientos de NA puede aumentar la mortalidad[5,26,27]. En la cohorte SU se observó que un intervalo más largo en el inicio del segundo vasopresor se asoció significativamente con la mortalidad hospitalaria, particularmente en los pacientes identificados de alto riesgo con la ESRR. Además, el momento de la administración de corticosteroides en el SS podría influir en la evolución de los pacientes[28,29]. La SSC recomienda iniciar vasopresina cuando la dosis de NA esté en el intervalo de 0,25-0,5 µg/kg/min, y administrar corticosteroides si la dosis de NA o adrenalina es ≥ 0,25 mcg/kg/min a las 4 horas de su inicio[3]. Por lo tanto, la puntuación en la ESSR podría ser útil para identificar a los pacientes que requieran vasopresores y corticosteroides.

Se empleó un algoritmo basado en aprendizaje automático, en concreto *random forest,* para seleccionar las variables predictoras. Esta metodología ofrece varias ventajas. Se trata de una herramienta más flexible y potente para desarrollar modelos de riesgo complejos en diversos ámbitos, y logra una mayor precisión predictiva. Sin embargo, el aprendizaje automático es difícil de adaptar al entorno de la práctica clínica real en términos de sistemas de historias clínicas electrónicas y nivel de explicación[10,30,31]. Por otra parte, los modelos predictivos tradicionales son sencillos y fácilmente interpretables, pero su capacidad para manejar un amplio conjunto de variables es limitada. *AutoScore* se diseñó para crear modelos predictivos que puedan implementarse y validarse fácilmente mediante aprendizaje automático y de regresión. Se ha utilizado para crear escalas[10,11,13,32]. El módulo de ajuste de *AutoScore* podría optimizar el

Tabla 3. Capacidad de la escala para predecir *shock* séptico refractario

	Punto de decisión en la ESSR	Sensibilidad (IC 95%)	Especificidad (IC 95%)	VPP (IC 95%)	VPN (IC 95%)
Cohorte MIMIC-IV (validación interna)	≥ 6 puntos	0,912 (0,873-0,946)	0,586 (0,574-0,599)	0,073 (0,069-0,076)	0,995 (0,992-0,997)
	≥ 12 puntos	0,686 (0,618-0,750)	0,903 (0,896-0,911)	0,200 (0,181-0,220)	0,988 (0,985-0,990)
Cohorte SU (a la llegada)	≥ 6 puntos	0,981 (0,966-0,993)	0,110 (0,099-0,123)	0,144 (0,142-0,147)	0,974 (0,956-0,990)
	≥ 12 puntos	0,739 (0,695-0,782)	0,533 (0,514-0,552)	0,195 (0,184-0,206)	0,930 (0,919-0,941)
Cohorte SU (al diagnóstico de la hipoperfusión o de la hipotensión)	≥ 6 puntos	0,986 (0,973-0,995)	0,120 (0,107-0,132)	0,146 (0,144-0,148)	0,982 (0,966-0,994)
	≥ 12 puntos	0,799 (0,760-0,835)	0,617 (0,599-0,636)	0,242 (0,230-0,255)	0,953 (0,944-0,961)
Cohorte UCI	≥ 6 puntos	0,941 (0,905-0,976)	0,379 (0,346-0,412)	0,231 (0,220-0,243)	0,970 (0,951-0,986)
	≥ 12 puntos	0,691 (0,619-0,756)	0,808 (0,780-0,834)	0,416 (0,375-0,457)	0,929 (0,914-0,944)

UCI: unidad de cuidados intensivos; SU: servicio de urgencias; VPN: valor predictivo negativo; VPP: valor predictivo positivo; ESSR: Escala de *shock* séptico refractario.

Tabla 4. Incidencia de eventos en función del grupo de riesgo de la Escala de *shock* séptico refractario

	Eventos	Riesgo bajo (ESSR < 6 puntos)	Riesgo intermedio (ESSR 6-11 puntos)	Riesgo alto (ESSR ≥ 12 puntos)
Cohorte MIMIC-IV (validación interna) (%)	*Shock* séptico refractario	0,5	2,5	20,1
	Dosis equivalente de noradrenalina ≥ 0,25 µg/kg/min	2,2	7,4	37,8
	Mortalidad intrahospitalaria	6,1	13,7	37,5
Cohorte SU (a la llegada) (%)	*Shock* séptico refractario	2,6	8,1	19,5
	Dosis equivalente de noradrenalina ≥ 0,25 µg/kg/min	14,7	24,0	42,8
	Mortalidad intrahospitalaria	12,1	15,5	26,9
Cohorte SU (al diagnóstico de la hipoperfusión o de la hipotensión) (%)	*Shock* séptico refractario	1,8	5,4	24,2
	Dosis equivalente de noradrenalina ≥ 0,25 µg/kg/min	6,7	19,6	52,2
	Mortalidad intrahospitalaria	10,0	14,3	30,4
Cohorte UCI (%)	*Shock* séptico refractario	3,0	10,0	41,6
	Dosis equivalente de noradrenalina ≥ 0,25 µg/kg/min	14,2	25,0	64,5
	Mortalidad intrahospitalaria	17,8	25,7	55,2

UCI: unidad de cuidados intensivos; SU: servicio de urgencias; VPN: valor predictivo negativo; VPP: valor predictivo positivo; ESSR: Escala de *shock* séptico refractario.

modelo para garantizar su relevancia clínica y su coherencia con patrones no lineales.

La ESSR incluye variables clínicas: presión arterial, frecuencia cardiaca, temperatura, lactato, pH arterial y cifra de leucocitos, que están asociadas con la gravedad del SS[25,33-36]. Estas variables son fáciles de determinar en la valoración inicial de una estancia en el SU o en la UCI. La escala se diseñó para representar las variables de forma ponderada, imitando la intuición del médico en la práctica clínica.

En el desarrollo de la ESSR se utilizó una gran base de datos de pacientes ingresados en UCI y se validó en otros entornos, incluidos los SU. Esta es una de las fortalezas del trabajo. Sin embargo, la heterogeneidad de las diferentes cohortes puede influir en la capacidad predictiva de la escala. La base MIMIC-IV incluyó pacientes con sospecha de infección en todos los tipos de UCI, lo que implica la inclusión de pacientes menos graves y a una baja prevalencia de SSR. Por el contrario, la cohorte SU incluyó pacientes con SS que cumplían la definición de sepsis, mientras que la cohorte UCI incluyó pacientes con sepsis procedentes principalmente de UCI de enfermedades médicas, lo que se tradujo en una mayor gravedad.

La capacidad predictiva fue más limitada en la inclusión del paciente en la valoración inicial en el SU. Sin embargo, mejora si se utiliza cuando se objetiva la hipoperfusión o la hipotensión en el SU. Por lo tanto, la escala puede tener la ventaja de predecir el SSR en diferentes momentos de la atención del paciente en el SU, identificando los cambios en el estado del paciente.

Este estudio tiene limitaciones. En primer lugar, el diseño retrospectivo puede haber generado sesgos. En segundo lugar, se utilizaron datos existentes y no se realizó un cálculo del tamaño de la muestra adaptado a nuestro modelo. Dado que se basó en aprendizaje automático y la proporción de eventos era pequeña, el tamaño de la muestra podría ser inadecuado. En tercer lugar, es posible que existan otras variables asociadas al pronóstico, como los niveles de péptido natriurético cerebral (BNP) o los hallazgos ecocardiográficos a pie de cama que no se han podido analizar por no estar disponibles. En cuarto lugar, la variabilidad en la práctica clínica entre centros puede afectar a la generalización de los resultados.

En conclusión, la ESSR, escala desarrollada con un enfoque híbrido aprendizaje automático-regresión, es capaz de identificar con una buena capacidad predictiva a los pacientes en riesgo de SSR en la UCI y en el SU. La ESSR puede utilizarse para la estratificación rápida del riesgo de los pacientes con SS, lo que permitiría iniciar precozmente un manejo intensivo en los pacientes con mayor riesgo de progresar a SSR. Se necesitan estudios prospectivos para validar y confirmar la aplicabilidad clínica de esta escala.

Conflicto de intereses: Los autores declaran no tener conflictos de interés en relación con el presente artículo.

Financiación: Este estudio fue financiado de forma parcial por una beca del National Basic Science Research Program a través de National Research Foundation of Korea (NRF) financiado por el Ministerio de Ciencia y TIC (RS-2023-00251786) y Samsung Medical Center (beca No. SMX1240281) a Tae Gun Shin.

Responsabilidades éticas: Todos los autores han confirmado el mantenimiento de la confidencialidad y respeto de los derechos de los pacientes en el documento de responsabilidades del autor, acuerdo de publicación y cesión de derechos a EMERGENCIAS.

Artículo no encargado por el Comité Editorial y con revisión externa por pares.

Bibliografía

1 Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315:801-10.
2 Namgung M, Ahn C, Park Y, Kwak IY, Lee J, Won M. Mortality among adult patients with sepsis and septic shock in Korea: a systematic review and meta-analysis. Clin Exp Emerg Med. 2023;10:157-71.
3 Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med. 2021;47:1181-247.
4 Suh GJ, Shin TG, Kwon WY, Kim K, Jo YH, Choi SH, et al. Hemodynamic management of septic shock: beyond the Surviving Sepsis Campaign guidelines. Clin Exp Emerg Med. 2023;10:255-64.
5 Teja B, Bosch NA, Walkey AJ. How We Escalate Vasopressor and Corticosteroid Therapy in Patients With Septic Shock. Chest. 2023;163:567-74.
6 Antonucci E, Polo T, Giovini M, Girardis M, Martin-Loeches I, Nielsen ND, et al. Refractory septic shock and alternative wordings: A systematic review of literature. J Crit Care. 2023;75:154258.
7 Nandhabalan P, Ioannou N, Meadows C, Wyncoll D. Refractory septic shock: our pragmatic approach. Crit Care. 2018;22:215.

8 Jentzer JC, Vallabhajosyula S, Khanna AK, Chawla LS, Busse LW, Kashani KB. Management of Refractory Vasodilatory Shock. Chest. 2018;154:416-26.
9 Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med. 2021;47:1181-247.
10 Xie F, Chakraborty B, Ong MEH, Goldstein BA, Liu N. AutoScore: A Machine Learning-Based Automatic Clinical Score Generator and Its Application to Mortality Prediction Using Electronic Health Records. JMIR Med Inform. 2020;8:e21798.
11 Xie F, Ong MEH, Liew J, Tan KBK, Ho AFW, Nadarajan GD, et al. Development and Assessment of an Interpretable Machine Learning Triage Tool for Estimating Mortality After Emergency Admissions. JAMA Netw Open. 2021;4:e2118467.
12 Liu N, Liu M, Chen X, Ning Y, Lee JW, Siddiqui FJ, et al. Development and validation of an interpretable prehospital return of spontaneous circulation (P-ROSC) score for patients with out-of-hospital cardiac arrest using machine learning: A retrospective study. EClinicalMedicine. 2022;48:101422.
13 Yu JY, Heo S, Xie F, Liu N, Yoon SY, Chang HS, et al. Development and Asian-wide validation of the Grade for Interpretable Field Triage (GIFT) for predicting mortality in pre-hospital patients using the Pan-Asian Trauma Outcomes Study (PATOS). Lancet Reg Health West Pac. 2023;34:100733.
14 Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10:1.
15 Hwang S-Y, Kim I-K, Jeong D, Park JE, Lee GT, Yoo J, et al. Prognostic Performance of Sequential Organ Failure Assessment, Acute Physiology and Chronic Health Evaluation III, and Simplified Acute Physiology Score II Scores in Patients with Suspected Infection According to Intensive Care Unit Type. J Clin Med. 2023;12:6402.
16 Lee HJ, Ko BS, Ryoo SM, Han E, Suh GJ, Choi SH, et al. Modified cardiovascular SOFA score in sepsis: development and internal and external validation. BMC Med. 2022;20:263.
17 You JS, Park YS, Chung SP, Lee HS, Jeon S, Kim WY, et al. Relationship between time of emergency department admission and adherence to the Surviving Sepsis Campaign bundle in patients with septic shock. Crit Care. 2022;26:43.
18 Ko BS, Ryoo SM, Han E, Chang H, Yune CJ, Lee HJ, et al. Modified Cardiovascular Sequential Organ Failure Assessment Score in Sepsis: External Validation in Intensive Care Unit Patients. J Korean Med Sci. 2023;38:e418.
19 Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med. 1996;22:707-10.
20 Johnson AE, Stone DJ, Celi LA, Pollard TJ. The MIMIC Code Repository: enabling reproducibility in critical care research. J Am Med Inform Assoc. 2018;25:32-9.
21 Jenkins CR, Gomersall CD, Leung P, Joynt GM. Outcome of patients receiving high dose vasopressor therapy: a retrospective cohort study. Anaesth Intensive Care. 2009;37:286-9.
22 Bassi E, Park M, Azevedo LC. Therapeutic strategies for high-dose vasopressor-dependent shock. Crit Care Res Pract. 2013;2013:654708.
23 Lambden S, Laterre PF, Levy MM, Francois B. The SOFA score-development, utility and challenges of accurate assessment in clinical trials. Crit Care. 2019;23:374.
24 Xie F, Ning Y, Liu M, Li S, Saffari SE, Yuan H, et al. A universal AutoScore framework to develop interpretable scoring systems for predicting common types of clinical outcomes. STAR Protoc. 2023;4:102302.
25 Conrad M, Perez P, Thivilier C, Levy B. Early prediction of norepinephrine dependency and refractory septic shock with a multimodal approach of vascular failure. J Crit Care. 2015;30:739-43.
26 Brask AL, Shemanski SM, Barnes TE, Holmes AK. Timing of Vasopressin Addition to Norepinephrine and Efficacy Outcomes in Patients With Septic Shock. Ann Pharmacother. 2023;57:521-6.
27 Sacha GL, Lam SW, Wang L, Duggal A, Reddy AJ, Bauer SR. Association of Catecholamine Dose, Lactate, and Shock Duration at Vasopressin Initiation With Mortality in Patients With Septic Shock. Crit Care Med. 2022;50:614-23.
28 Zhang L, Gu WJ, Huang T, Lyu J, Yin H. The Timing of Initiating Hydrocortisone and Long-term Mortality in Septic Shock. Anesth Analg. 2023;137:850-8.
29 Sacha GL, Chen AY, Palm NM, Duggal A. Evaluation of the Initiation Timing of Hydrocortisone in Adult Patients With Septic Shock. Shock. 2021;55:488-94.
30 Kang CY, Yoon JH. Current challenges in adopting machine learning to critical care and emergency medicine. Clin Exp Emerg Med. 2023;10:132-7.
31 Okada Y, Ning Y, Ong MEH. Explainable artificial intelligence in emergency medicine: an overview. Clin Exp Emerg Med. 2023;10:354-62.
32 Xie F, Liu N, Yan L, Ning Y, Lim KK, Gong C, et al. Development and validation of an interpretable machine learning scoring tool for estimating time to emergency readmissions. EClinicalMedicine. 2022;45:101315.
33 Capp R, Horton CL, Takhar SS, Ginde AA, Peak DA, Zane R, et al. Predictors of patients who present to the emergency department with sepsis and progress to septic shock between 4 and 48 hours of emergency department arrival. Crit Care Med. 2015;43:983-8.
34 Noritomi DT, Soriano FG, Kellum JA, Cappi SB, Biselli PJ, Libório AB, et al. Metabolic acidosis in patients with severe sepsis and septic shock: a longitudinal quantitative study. Crit Care Med. 2009;37:2733-9.
35 Leisman DE, Doerfler ME, Schneider SM, Masick KD, D'Amore JA, D'Angelo JK. Predictors, Prevalence, and Outcomes of Early Crystalloid Responsiveness Among Initially Hypotensive Patients With Sepsis and Septic Shock&ast. Crit Care Med. 2018;46:189-98.
36 Belok SH, Bosch NA, Klings ES, Walkey AJ. Evaluation of leukopenia during sepsis as a marker of sepsis-defining organ dysfunction. PLoS One. 2021;16:e0252206.